

作文の評価手順が評価に及ぼす影響について —analytic scoring の採点に関して—

三谷閑子

名古屋大学大学院国際言語文化研究科

silvernuttree@yahoo.co.jp

1. はじめに

2002 年から実施され始めた日本留学試験には、従来の日本語能力試験にはなかった「記述問題」が導入されている。日本の大学への留学を希望する受験生が、大学の教養課程や専門課程で要求される文章を書く力をどの程度持っているかを測ることは重要であり、このような記述問題が導入されるようになったことには非常に意義があると言える。しかし、その評価方法には、評価基準を含め、まだ検討を要する点があると考えられる。

そこで、評価結果をより活かせるような新しい評価基準¹を作成する研究会を開いているが、新基準の信頼性を検討していく過程で、評価者間の採点手順の違いが評価の結果に影響を及ぼしているのではないかという疑問が出てきた。本発表は、この疑問に基づき実際に調査を行った結果の報告である。本研究は、最終的には評価の信頼性と妥当性、及び評価者トレーニングへの示唆を得ることを目的とする。

本研究では **analytic scoring** の際の採点手順を評価者のストラテジーの 1 つとして捉える。作文の評価方法にはおおまかに分けて **holistic scoring** と **analytic scoring** の 2 つ² があるが、本研究では特に **analytic scoring** の場合の評価過程に焦点を当てる。

2. 先行研究

- 評価者の評価ストラテジー³ ⇒ 評価を決定するまでの思考と行動過程、作文の読み方、判断を下す時期など様々なものを含む。
- **holistic scorer** の評価過程
 - Cumming (1990)、Vaughan (1991)、Huot (1993)、Pula & Huot (1993)
 - 評価者の熟達度別評価ストラテジーの違い
 - Wolfe & Ranney (1996)、Wolfe (1997)
- **analytic scorer** の評価過程
 - DeRemer (1998) 評価者の作業過程が評価に及ぼす影響

Oholistic scoring と比べて評価者の自由をかなり制限した analytic scoring の場合であっても評価者間に異なるストラテジーが存在し、それが評価に影響する可能性があると考えられる。

3. 研究課題

I. analytic scoring の場合でも評価者間に評価ストラテジーの違いが認められるか。認められる場合、どのような違いが見られるか。

II. 採点手順は評価に影響するか。

① 採点手順の違いは評価の一致度、採点時間に影響を与えるか。

② 採点手順の違いは各評価項目の独立性に影響を与えるか。

III. 評価者の個人差と条件差のどちらが評価に大きく影響するか。

4. データ

学習者の作文：外国語使用についての意見文（平成 14 年度日本留学試験「記述問題」）

調査 I 20 人分 調査 II 66 人分

5. 研究方法

5-1. 評価基準

7つの観点項目： ①正確さ ②文体の統一 ③語彙の多様性 ④文のわかりにくさ ⑤文間
⑥段落間 ⑦内容

5-2. 採点手順

① 個人別採点方式： 作文ごとに全観点項目を採点するやり方。

② 項目別採点方式： 観点項目ごとに全作文を採点していくやり方。

③ 部分的項目別採点方式：

7つの観点項目のうち、「正確さ」「文体」「語彙の多様性」は項目別採点方式で採点し、残りの観点項目はまとめて一度に行うやり方。

5-3. 調査 1（研究課題 I）

データを 2 等分し、10 枚を項目別採点方式、残りの 10 枚を個人別採点方式で採点した。採点後、評価者が集まって採点中に気づいたことを話し合い、テープに録音した。

5-4. 調査2（研究課題Ⅱ、Ⅲ）

データを3等分し、それぞれ、項目別採点方式、個人別採点方式、部分的項目別採点方式の手順で採点し、評価者間の評価の一致度、採点時間、評価観点項目の項目間相関を比較した。次に、項目別採点方式と個人別採点方式のデータを対象にして、個人差と条件差のどちらがより評価に影響を及ぼしているかを、二元配置の分散分析の手法を使って調べた。

6. 結果

6.1 研究課題Ⅰ（調査1）

- ・個人別採点方式の場合には、採点項目の順番にかなり大きな個人差が認められた。評価者全員に共通の特徴としては、意味に関わる観点項目、正確さに関わる観点項目をそれぞれまとめて同時に採点する傾向がある。
- ・採点時間： 項目別採点方式 > 個人別採点方式

6.2 研究課題Ⅱ（調査2）

- ① ・3つの採点方式と評価者の一致度： 3つにはあまり差はなく（ α 係数：項目別=0.88、個人別=0.93、部分的項目別採点方式=0.89）、比較的高い一致度が得られた。
 - ・採点時間： 項目別採点方式 > 部分的項目別採点方式 > 個人別採点方式
- ② ・項目間相関： 項目別採点方式 > 部分的項目別採点方式 > 個人別採点方式
→項目別採点方式は各項目の独立性が強い。

6.3 研究課題Ⅲ（調査2）

- ・条件差（採点方式の違い）と個人差両方が評価結果に有意な差を生み出している。しかし、両者を比較した場合、全体的に見ると条件差の方が個人差よりもやや強く評価に影響を及ぼしていると考えられる。

7. 結論と示唆

採点手順の違いは評価の信頼性に強い影響を及ぼしていなかったが、評価の妥当性と時間的効率の観点からは、採点手順の種類を慎重に選んだほうがよいと言える。この点で、本研究で試みた部分的個別採点方式は実際に評価を行っていくときに有効な採点方式の一つであろうと考えられる。また、評価者の自由度が制限される **analytic scoring** の場合であっても、やはり評価者間にストラテジーの違いがあることが確認された。これは、**analytic scoring** の

場合も **holistic scoring** と同様に評価者トレーニングが必要であることを示唆している。

8. 今後の課題

- ・再検証
- ・評価者の内省を **think aloud** 法を使って調査

<参考文献>

- ・ Cumming, A. 1990. Expertise in evaluating second language compositions. *Language Testing*, 7, 31-51.
- ・ DeRemer, M. 1998. Writing assessment: rater's elaboration of the rating task. *Assessing Writing*, 5(1), 7-29.
- ・ Huot, B. 1993. The influence of holistic scoring procedures on reading and rating student writing. In M. Williamson & B. Huot (Eds.), *Validating holistic scoring for writing assessment* (pp.206-236). Cresskill, NJ: Hampton Press Inc.
- ・ Pula, J., & Huot, B. 1993. A model of background influences on holistic raters. In M. Williamson & B. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp.237-265). Cresskill, NJ: Hampton Press Inc.
- ・ Vaughan, C. 1991. Holistic assessment: what goes on in the rater's mind? In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp.111-125). Westport, NJ: Ablex.
- ・ Wolfe, E.W., & M. Ranney. 1996. Expertise in essay scoring. In D. C. Edelson & Domeshek (Eds.), *Proceedings of ICLS 96* (pp.545-550). Charlottesville, VA: Association for the Advancement of Computing in Education.
- ・ Wolfe, E.W. 1997. The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, 4(1), 83-106.

¹ 名古屋大学 平成 15、16 年度科研費研究「日本留学試験における記述問題の実施方法と分析観点に関する実証的研究—記述問題の問題形式・量および評価基準の適正さについて」

² **holistic scoring** とは、作文の全体的印象を基に単一の得点で作文の評価をする評価方法のことである。Educational Testing Service が行う TOEFL のライティング・テストや British Council が行う ELTS はこの方法を採用している。この評価方法に対して、**analytic scoring** は作文を複数の観点から採点し、最後に個々の観点の得点を合計して評価する。

³ 評価ストラテジーを評価プロセスと呼ぶ論文 (Lumley 2002) もあるが、本稿では評価ストラテジーとして扱う。先行研究では、評価プロセスと評価ストラテジーははっきり区別されていない。